DOI:
10.1002/aur.1466

Document Version
Publisher's PDF, also known as Version of record

New Interview and Observation Measures of the Broader Autism Phenotype: Description of Strategy and Reliability Findings for the Interview Measures

Jeremy R. Parr, Maretha V. De Jonge, Simon Wallace, Andrew Pickles, Michael L. Rutter, Ann S. Le Couteur, Herman van Engeland, Kerstin Wittemeyer, Helen McConachie, Bernadette Roge, Carine Mantoulan, Lennart Pedersen, Torben Isager, Fritz Poustka, Sven Bolte, Patrick Bolton, Emma Weisblatt, Jonathan Green, Katerina Papanikolaou, Gillian Baird, and Anthony J. Bailey

Clinical genetic studies confirm the broader autism phenotype (BAP) in some relatives of individuals with autism, but there are few standardized assessment measures. We developed three BAP measures (informant interview, self-report interview, and impression of interviewee observational scale) and describe the development strategy and findings from the interviews. International Molecular Genetic Study of Autism Consortium data were collected from families containing at least two individuals with autism. Comparison of the informant and self-report interviews was restricted to samples in which the interviews were undertaken by different researchers from that site (251 UK informants, 119 from the Netherlands). Researchers produced vignettes that were rated blind by others. Retest reliability was assessed in 45 participants. Agreement between live scoring and vignette ratings was very high. Retest stability for the interviews was high. Factor analysis indicated a first factor comprising social-communication items and rigidity (but not other repetitive domain items), and a second factor comprised mainly of reading and spelling impairments. Whole scale Cronbach's alphas were high for both interviews. The correlation between interviews for factor 1 was moderate (adult items 0.50; childhood items 0.43); Kappa values for between-interview agreement on individual items were mainly low. The correlations between individual items and total score were moderate. The inclusion of several factor 2 items lowered the overall Cronbach's alpha for the total set. Both interview measures showed good reliability and substantial stability over time, but the findings were better for factor 1 than factor 2. We recommend factor 1 scores be used for characterising the BAP. ***Autism Res*** *2015, 8: 522–533.* © 2015 The Authors Autism Research published by Wiley Periodicals, Inc. on behalf of International Society for Autism Research

Keywords: broader autism phenotype; informant interview; self-report interview; interrater reliability; retest reliability

From at least the time of the first systematic epidemiological study of autism by Lotter (1967), it was apparent that there were many individuals who showed autism features that were either milder or fewer in number than those required for the traditional diagnostic cut-offs [Lotter, 1967]. The later Wing and Gould study (1979) showed much the same and it became clear that there were unresolved uncertainties over where and how to draw the diagnostic boundaries [Wing & Gould, 1979]. The importance of the issue was highlighted by Folstein & Rutter's [1977a] twin study findings, followed by the expansion of the twin sample [Bailey et al., 1995; Le Couteur et al., 1996], showing that the underlying genetic liability spanned both definite autism spectrum disorder (ASD) diagnoses and qualitatively similar, but milder, socio-communicative deficits. These findings and those from Folstein and her colleagues [Landa et al., 1992; Piven, Palmer, Jacobi, Childress, & Arndt, 1997] and

From University of Oxford Department of Psychiatry, UK (J.R.P., S.W., K.W., A.J.B.); Institutes of Neuroscience, and Health and Society, Newcastle University, UK (J.R.P., A.S.L.C., H.M.); Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands (M.V.D.J., H.V.E.); Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK (A.P., P.B.); MRC Centre for Social, Genetic and Developmental Psychiatry, King's College London, UK (M.L.R.); Centre d'Etudes et de Recherches en Psychopathologie, Toulouse, France (B.R., C.M.); Center for Autisme, Denmark (L.P., T.I.); Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, Germany (F.P., S.B.); Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden (S.B.); Cambridge University and Department of General and Adolescent Paediatrics, Institute of Child Health, University College London, London, UK (E.W.); Academic Department of Child Psychiatry, University of Manchester, Manchester, UK (J.G.); University Department of Child Psychiatry, Athens, Greece (K.P.); Guy's and St Thomas' NHS Trust & King's College London, UK (G.B.); Department of Psychiatry, University of British Columbia, Vancouver, Canada (A.J.B.)

Received June 04, 2013; accepted for publication January 31, 2015

The FHI-I and FHI-S are available from the corresponding author on request.

Address for correspondence and reprints: Jeremy Parr, Institute of Neuroscience, Newcastle University, UK. E-mail: Jeremy.Parr@ncl.ac.uk.

Published online 10 May 2015 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.1466

from Szatmari et al., [2000] gave rise to the concept of a broader autism phenotype (BAP) [Losh, Adolphs, & Piven, 2011]. Family study findings [Bolton et al., 1994] showed that in parents and siblings these broader phenotype features were more common in families containing an individual with autism than families of an individual with Down syndrome (DS).

In parallel with this work, evidence accumulated to show that both causal risk factors and disorders themselves operated dimensionally [Rutter, 2003, 2009]. This finding applied across the whole of medicine with respect to multifactorial disorders and not just to neuropsychiatric conditions. In psychopathology, this was obviously the case with depression and conduct disorders, but there were many indications that it might apply to both autism and schizophrenia—giving rise in both cases to the notion of a spectrum. At much the same time, behavioral geneticists were pointing out that the same genetic principles operated with respect to the liability to categorical disorders and to dimensions—the latter being influenced by quantitative trait loci (QTLs) [Plomin, Haworth, & Davis, 2009]. That recognition opened the way to expanding molecular genetic strategies to include QTL approaches.

During the 1990s, reliable and valid standardized interviews—Autism Diagnostic Interview [Rutter, Le Couteur, & Lord, 2003] and observation methods—Autism Diagnostic Observational Schedule [Lord, Rutter, DiLavore, & Risi, 2001] were developed for the categorical diagnosis of autism. Subsequently these instruments were widely taken up and became accepted as the "gold standards" to assist clinicians in ASD diagnosis. However, ASD measures were designed to identify the extent to which an individual had diagnostic features of ASD, and met criteria for diagnosis, rather than having some more subtle features of the BAP. Specific BAP measures were, therefore, produced. These included an informant family history interview (on which there was no systematic assessment of validity) [Bolton et al., 1994], and a rudimentary impression of informant measure used to rate interviewers observations about an interviewee that lacking systematic sample coverage, was never reported. Subsequently, Pickles et al. concluded that the identification of the BAP required the use of multiple measurement methods, including self and informant report, and direct observation of the relevant features, to "triangulate" the phenotype, and not rely on one source of data [Pickles et al., 2000]. Similar views were expressed by Dawson et al. [2002]; Bailey and colleagues summarized the main findings up to that date [Bailey, Palferman, Heavey, & Le Couteur, 1998].

Research Strategy, Including the Development of New Measures

In 2000–2001, the International Molecular Genetic Study of Autism Consortium (IMGSAC) constructed new measures to dimensionalize the individual components of the BAP, in such a way that allowed their use with individuals with and without traditionally diagnozed ASD. First, new informant and self-report versions of the Family History Interviews were developed. The content was indicated by many items included in the original family history schedule [Bolton et al., 1994], and the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) ratings, accepting the need to capture more subtle yet qualitatively similar ASD behavioral characteristics. Substantial changes were needed to conceptualize, identify, and rate the rigid, repetitive, and stereotyped aspects of activities and behaviors that might be associated with broader ASD. In addition, some new items were added to the interviews, for example, to cover pragmatic and conversational qualities, emotional intimacy, demonstrativeness, and response to emotional cues. Finally, decisions were made to retain some items on neurodevelopmental and mental health disorders, (reading and spelling difficulties, anxiety disorder, anxious worrying, depression, and bipolar disorder) and exclude other items (such as dementia, tics, and immune disorders) thought unlikely to be helpful in defining the BAP.

Alongside the family history interviews, a new direct observation BAP measure (the impression of interviewee interview [IoI] was created, based on clinical research experience and the limited information available from the published literature, for example, the Pragmatic Rating Scale [Landa, Wzorek, Piven, Folstein, & Isaacs, 1991], and the observational items in the ADOS [Lord et al., 2001].

Items were included that focused on qualitatively similar social/communicative deficits and a range of other related behaviors reported in the autism and developmental disability literature, the international classification systems (ICD/DSM), the limited published research findings on BAP and clinical practice. A decision was made to include a relatively broad range of items across the domains of interest while acknowledging the uncertainty about the relevance of certain aspects of early development to the conceptual framework of the BAP.

IMGSACs measures, therefore, included an informant Family History Interview (FHI-I), a self-report subject interview (FHI-S), and researcher observation ratings: the Impression of the interviewee—IoI. The adult FHI-I and FHI-S contain 77 items and take around 30–60 min to complete for a trained researcher, with the duration dependent on the extent of the elicited BAP traits and behaviors. The children's FHI-I and FHI-S take 20–30 min to complete. Mandatory and optional probes allowed the interviewer to gain examples of behavior; subsequently if insufficient information to score was

EMOTIONAL CUES AND RESPONSIVENESS IN CHILDHOOD ***Focus on age 5.0-15.11.***

This item focuses on the subject's sensitivity to the expressed emotions of others and their response.

As a child how were you at picking up on how other people were feeling? For instance if someone was upset could you tell? Did you ever have difficulty recognising whether people were happy, sad or angry?

Did you ever have difficulty responding to whether people were happy, sad or angry? How did you usually react when someone was obviously sad? Did you usually try to comfort them?

If difficulties mentioned, probe:

Did you ever get into trouble for not noticing or responding to how someone was feeling?

0 = no difficulties in recognising and responding to emotional cues

1 = some difficulties recognising and responding to emotional cues; may have been remarked upon by others, but not associated with significant social difficulties

2 = persistent and significant difficulties recognising and responding to emotional cues; code here if subject's behaviour led to distress, conflicts or avoidance of emotionally charged situations, or was associated with teasing/bullying

Figure 1. The emotional cues and responsiveness item from the FHI-S childhood section.

not available, the interviewer asked their own supplemental questions to gain examples of behavior to allow scoring (an example of an FHI-S item is given in Fig. 1). Behaviors were scored as "0" (behavior does not reach scoring threshold); "1" (difficulties of the type specified but not associated with impairment); or "2" (associated impairment). Separate versions of the interviews suitable for children were created; the conceptual content was identical to that of the adult version although minor, but important changes, in wording (and/or sequence) were required. The aim was that these three BAP measures triangulated the BAP of individuals, reducing the influence of lack of insight of interviewees, or bias. This article describes the psychometric properties and reliability of the FHI-I and FHI-S. The second paper in this series [Pickles et al., 2013] focuses on the properties and reliability of the IoI. The third paper includes data from parents of children with ASD, and parents of children with DS, and presents the findings on the discriminative validity of the measures and draws conclusions on the utility of the trio of measures [de Jonge et al., 2014].

Methods and Materials

Ethical approval for the study was obtained in each country and all participants gave written informed consent. IMGSAC families with at least two individuals with ASD were identified in participating countries (IMGSAC 1998, IMGSAC 2001). Principal investigators (PIs) and researchers were trained on FHI-I and FHI-S administration. For UK and Dutch families, two or more researchers visited each family to carry out independent FHI-I and FHI-S. In other countries, it was often necessary for both interviews to be administered by one researcher. In all cases, the researcher administering the subject interview (FHI-S) scored the items on the IoI. For adults, the FHIs had two sets of items, one that related to childhood and one to adulthood, while for non-ASD children up to age 16 years, there was a single item set. Children aged 11 years and above were interviewed using the children's FHI-S. Mothers reported about non ASD siblings aged 4 and above.

Reliability Testing

Items scores used for reliability were derived in two ways: "live" codings made directly by the interviewer (termed "live ratings") and consensus codings of anonymized vignettes, written following all interviews, by the researcher who undertook the FHI-S and IoI and the researcher who undertook the FHI-I. The anonymized vignettes were up to three pages long, and contained all relevant information (behavioral/characteristic examples) required for coding, together with gender, and approximate age but omitted any details that might indicate group (the data from de Jonge's DS parents were also scored during this process) or family type (cousin pairs or sibling pairs) [de Jonge et al., 2014]. Following joint training of PIs and researchers, vignettes were scored separately by at least one

Table 1. Percentage of Males and Females Scoring "1"/"2"[a] for Each Item

	FHI-I (childhood)		FHI-I (adulthood)		FHI-S (childhood)		FHI-S (adulthood)	
Item	Males ($n = 132$)	Females ($n = 138$)	Males ($n = 154$)	Females ($n = 148$)	Males ($n = 157$)	Females ($n = 159$)	Males ($n = 157$)	Females ($n = 161$)
Delay in spoken language	1%/0%[a]	1%/2%	NA	NA	4%/3%	0%/4%	NA	NA
Articulation	2/0	2/0	NA	NA	5/0	3/0	NA	NA
Reading	5/2	3/1	6/1	3/1	12/4	13/4	6/2	2/2
Spelling	14/9	10/5	**19/17**	**11/8**	**19/21**	**8/13**	**19/19**	**9/11**
Lack of interest in conversation	10/7	12/0	**28/10**	**10/2**	**31/7**	**8/2**	**28/3**	**11/3**
Reciprocal quality of conversation	**11/2**	**4/0**	**38/12**	**15/2**	**23/5**	**11/1**	**31/5**	**16/0**
Pedantic speech (childhood)	4/0	4/1	NA	NA	19/5	13/2	NA	NA
Pragmatics (adulthood)	NA	NA	25/8	17/4	NA	NA	**50/8**	**30/2**
Social play (childhood)	11/4	10/0	NA	NA	**17/4**	**7/2**	NA	NA
Aloof	**11/8**	**9/1**	**24/13**	**14/2**	21/3	15/1	**24/5**	**12/3**
Shyness	17/4	18/9	19/6	26/4	44/21	50/25	21/9	29/8
Friendships	**20/7**	**7/2**	**50/22**	**25/4**	**31/7**	**27/2**	**43/15**	**25/4**
Affection	14/4	14/1	**33/15**	**23/3**	29/12	23/5	32/3	16/6
Intimacy	NA	NA	**33/20**	**18/5**	NA	NA	**24/7**	**11/4**
Emotional cues and responsiveness	**22/0**	**9/0**	**49/21**	**18/1**	**49/9**	**17/3**	**30/6**	**11/1**
Demonstrativeness	15/4	17/1	**43/25**	**33/7**	30/8	30/4	**40/11**	**32/2**
Social behavior	6/0	6/1	**34/10**	**11/5**	**17/9**	**11/2**	**32/12**	**16/0**
Intensity of hobby	32/5	13/4	**40/25**	**27/7**	**65/13**	**47/6**	**47/21**	**36/6**
Social aspects of hobby	6/5	4/2	8/11	12/2	14/7	12/3	21/3	13/2
Circumscribed nature of hobby	8/2	3/2	8/5	8/0	**15/5**	**6/0**	18/3	8/2
Organisational skills	13/2	7/1	NA	NA	**22/2**	**7/2**	NA	NA
Rigid or perfectionistic in childhood	7/0	8/0	NA	NA	14/4	16/2	NA	NA
Rigidity (adulthood)	NA	NA	36/9	27/9	NA	NA	23/5	24/3
Perfectionism (adulthood)	NA	NA	39/4	32/3	NA	NA	35/8	34/3
Obsessive compulsive/ritualistic	2/0	4/0	9/5	6/4	10/1	9/3	14/4	17/4

The interview items that contribute to the factor 1 total scores are in bold text in the item list. NA = not applicable. Bold in the table cells = sex difference $P < 0.05$ from ordinal logistic Wald test.

[a] For each cell, data are presented as % scoring 1/% scoring 2

researcher from each collaborating site. When there was unanimous agreement on a coding, this was assigned by a researcher. Consensus coding agreements were achieved for the remaining items at regular consensus meetings attended by at least one member from all sites. This process was undertaken to ensure data quality and to maintain reliability. In the few cases, when medical or environmental factors might have had a major influence on a particular behavior, the codings for these behaviors were treated as inapplicable. The vignette approach was intended so that all family members were rated blind to family identity and collaborating site, but the researcher preparing the vignette could not be blind to family group or type. Several steps were taken to reduce potential biases in writing vignettes, including regular training and regular consensus reliability meetings. Nevertheless, the possibility of inadvertent bias through the consensus process could not be avoided completely, thus, reliability was also investigated through the live ratings, which had the advantage of being based on detailed descriptions of behaviors and face to face contact with the subject. However, the live ratings were not made blind to family type and group status, and therefore, we placed primary reliance on the consensus codes. We also assessed the agreement between the "live ratings" and consensus codes to inform future research use of the instruments; if the "live ratings" proved satisfactory, the FHIs would be more readily usable by other researchers by omitting the consensus coding stage.

Sample

We examine data from the IMGSAC families on 354 adults (parents and adult siblings, age 16–70 years), 61 siblings aged 4–15 years on the FHI-I and 385 adults, and 34 siblings on the FHI-S (age 11–15 years). Much of the analysis is restricted to the UK and Dutch samples for which subject and informant interviews conducted by different interviewers were available (UK 255 adult and 25 child informant, 251 adult and 28 child subject; Dutch 122 adult and 3 child informant, 119 adult and 3 child subject). No adults or siblings had received a clinical diagnosis of ASD.

Assessment of Reliability

Reliability was examined in three ways. First, the agreement between individual items and total scores were assessed as a measure of internal consistency for each interview. Second, the agreement between single "live ratings" and vignette consensus ratings were examined. Third, test–retest reliabilities were examined. Retest assessments were obtained on the FHI-I for 46 UK adults, and for the FHI-S on 45 UK adults reinterviewed 6–12 months after the original interviews by different interviewers blind to the original FHI-I, FHI-S, and IoI data. Anonymized vignettes were produced for the retest assessments and coded blind and separately by researchers from each collaborating site, and when necessary, using the study consensus coding procedures (see earlier).

Statistical Analysis

The factor analysis was undertaken in Mplus [Muthen and Muthen, 2008] using the unweighted least squares estimator for categorical data and we report the promax rotation. All other analyses were undertaken in Stata 11 [Statacorp, 2010].

Results

Internal Consistency and Test–Retest Reliability of Consensus Coded Items

Table 1 shows the frequency of the item codes for adult males and females. Table 2 shows item-test correlations and Cronbach alphas. Table 3 shows the item test–retest agreement. For the adult FHI-I, items from the childhood section showed moderate to good agreement, with most kappa values lying between 0.4 and 0.7. Agreement for items from the adulthood section about contemporaneous characteristics was greater, with 12/18 items showing good agreement (kappa > 0.60). FHI-S reliabilities were generally moderate for both childhood and adulthood items. For the FHI-I and FHI-S, agreement on cognitive, language, and socio-emotional items was higher than for repetitive domain items. For the FHI-S, children endorsed insufficient items to allow an adequately powered analysis.

Sixteen items (those listed in Table 5) showed reasonable frequency of occurrence, item-total correlations, and test-restest agreement across both subject and informant versions. Items on delay in spoken language, articulation, pedantic speech, social play, shyness, hobby items, and organisation skills were, therefore, excluded, and exploratory factor analyses of the remaining 16 adult items gave eigenvalues for males and females indicating two, three, or four factor solutions as plausible (Table 4). Large positive factor loading scores (Table 5) broadly consistent across genders and interview type suggested that a two factor solution appeared most meaningful. Eigenvalues and factor loading scores for childhood items were broadly similar to those found for adulthood items (Tables 4 and 5). For both the FHI-I and FHI-S, and for males and females, the first factor comprised a broad grouping of social-communication items and rigidity (but not other repetitive domain items); the second factor comprised reading and spelling impairments. For the 11 adulthood items, and the 9 childhood items that loaded on factor

Table 2. Adult FHI-I and FHI-S: Inter-Item Correlation Between Individual Items and Total Item Score, and Influence of Each Item on Cronbach Alpha ($n = 325$)

Items referring to childhood	FHI-I Item-test correlation between item and sum of items	FHI-S Item-test correlation between item and sum of items
Delay in spoken language	0.41	0.20
Articulation	0.37	0.27
Reading	0.25[a]	0.20
Spelling	0.30[a]	0.27[a]
Lack of interest in conversation	0.53	0.48
Reciprocal quality of conversation	0.66	0.37
Pedantic speech	0.35	0.27
Social play	0.67	0.57
Aloof	0.62	0.50
Shyness	0.21[a]	0.21[a]
Friendships	0.57	0.49
Affection	0.50	0.39
Emotional cues and responsiveness	0.73	0.51
Demonstrativeness	0.60	0.42
Social behavior	0.58	0.44
Intensity of hobby	0.51	0.28
Social aspects of hobby	0.50	0.43
Circumscribed nature of hobby	0.52	0.37
Organisational skills	0.34	0.33
Rigid or perfectionistic	0.40	0.42
Obsessive compulsive/ritualistic	0.32	0.22
Items referring to adulthood		
Reading	0.26	0.22
Spelling	0.25[a]	0.25[a]
Lack of interest in conversation	0.57	0.46
Reciprocal quality of conversation	0.62	0.33
Pragmatics	0.43	0.44
Aloof	0.58	0.52
Shyness	0.26	0.24[a]
Friendships	0.59	0.46
Affection	0.55	0.40
Intimacy	0.58	0.32
Emotional cues and responsiveness	0.67	0.53
Demonstrativeness	0.57	0.36
Social behavior	0.45	0.37
Intensity of hobby	0.36	0.29
Social aspects of hobby	0.25	0.27
Circumscribed nature of hobby	0.21	0.32
Rigidity/openness to experience	0.50	0.45
Perfectionism	0.21[a]	0.28
Obsessive compulsive/ritualistic	0.16[a]	0.31

[a] Item that reduces Cronbach alpha for whole scale.

1, the whole scale Cronbach alpha was high for both the FHI-I (0.870 and 0.845, respectively) and the FHI-S (0.781 and 0.758, respectively). The correlations between these childhood and adulthood item totals was 0.636 ($P < 0.001$) for the FHI-I and 0.664 ($P < 0.001$) for the FHI-S.

Comparison of Consensus FHI-I and FHI-S Scores

The comparison of parallel items from the childhood sections revealed high percentage agreements (range 72.5–98.9) between Subject and Informant interviews but variable kappa values (15 out of 28 values being between 0.20–0.39 and 4 items $\geq$ 0.40; Supporting Information Table A). For comparable adult items, percentage agreements were high (range 83.5–97.7) but again kappa values were generally low or moderate: (15 out of 25 kappa values were between 0.20 and 0.39 and 7 out of 25 items $\geq$ 0.40). Correlations were moderate between FHI-I and FHI-S for the factor 1 adult 11-item total score (0.528, $P < 0.0001$) and the childhood 9-item total (0.433, $P < 0.0001$); the interview items that contribute to the factor 1 total scores are in bold text in Table 1. The factor 1 FHI-I (parent report) total-score correlation with the FHI-S score from the 29 children who also completed subjects interviews was poor (0.089, $P = 0.646$), suggesting that the current version

Table 3. Test–Retest Reliability (Weighted Kappa Values) of Child and Adulthood Adult FHI-I and FHI-S

Interview item	FHI-I (childhood) Kappa ($n = 46$)	FHI-I (adulthood) Kappa ($n = 46$)	FHI-S (childhood) Kappa ($n = 45$)	FHI-S (adulthood) Kappa ($n = 45$)	Children's FHI-I Kappa ($n = 10$)	Children's FHI-S Kappa ($n = 6$)
Delay in spoken language	1.0	NA	0.72	NA	0.73	–
Articulation	–	NA	–	NA	–	–
Reading	0.60	0.38	0.84	0.85	0.68	0.86
Spelling	0.97	0.83	0.90	0.84	0.38	0.57
Lack of interest in conversation	0.46	0.73	0.52	0.35	–	0.57
Reciprocal quality of conversation	0.63	0.65	0.38	0.27	0.88	NA
Pedantic speech (childhood)	0.57	NA	0.52	NA	0.69	0.0
Pragmatics (adulthood)	NA	0.65	NA	0.47	NA	NA
Social play (childhood)	0.55	NA	0.49	NA	0.62	–
Aloof	0.79	0.55	0.32	0.47	–	–
Shyness	0.81	0.77	0.43	0.58	0.03	0.67
Friendships	0.66	0.50	0.44	0.46	0.76	NA
Affection	0.38	0.62	0.44	0.33	1.0	0.0
Intimacy	NA	0.60	NA	0.44	NA	NA
Emotional cues and responsiveness	0.40	0.66	0.42	0.91	0.91	-
Demonstrativeness	0.41	0.63	0.46	0.68	0.76	0.25
Social behavior	0.85	0.83	0.43	0.82	1.0	0.0
Extracurricular skills/ hobbies	–	–	–	–	–	–
Intensity of hobby	0.56	0.67	0.51	0.54	0.5	0.2
Social aspects of hobby	0.42	0.22	0.74	0.03	1.0	NA
Circumscribed nature of hobby	0.28	–	−0.10	−0.05	1.0	NA
Organisational skills	0.21	–	0.26	–	0.79	NA
Rigid or perfectionistic in childhood	0.52	–	−0.01	–	1.0	NA
Rigidity (adulthood)	NA	0.27	NA	0.76	NA	NA
Perfectionism (adulthood)	NA	0.68	NA	0.55	NA	NA
Obsessive compulsive/ritualistic	0.0	−0.05	0.39	0.75	–	0.6

NA indicates not applicable, indicates where positive score frequency was too low for analysis

of the children's FHI-S needs further investigation before it could be recommended for use.

The Behavioral Components of the BAP

Regarding consensus data, in adulthood, males were significantly more likely than females to show almost all BAP behaviors (Table 1). The rate of definite language delay self-reported was low (3% in females and males) but was more common in siblings (7.2% girls, 12.2% boys). For both the FHI-I and FHI-S, the factor1 item-total scores were significantly higher ($P < 0.001$) for men than women (FHI-I childhood 1.71 (2.94) vs. 0.70 (1.57); FHI-I adulthood 4.50 (4.63) vs. 1.69 (2.75); FHI-S childhood 2.44 (2.78) vs. 1.20 (1.95); FHI-S adulthood 3.22 (3.29) vs. 1.45 (2.21)). Adult siblings had similar factor1 item-total scores to parents according to the subject report (childhood FHI-I 1.83 (2.55) vs. FHI-S 1.74 (2.08) $P = 0.669$, adulthood FHI-I 1.62 (2.24) vs. FHI-S 2.41 (3.01) $P = 0.178$), but on the informant interview the adulthood total was lower (1.77 (3.36) vs. 3.25 (4.14) $P = 0.021$) while the childhood totals were not significantly different (2.19 (4.14) vs. 0.98 (1.93) $P = 0.247$).

Reliability of Live Scoring Method

Agreements between live and consensus scores were very high for the FHI-I and FHI-S (Supporting Information Table B), and kappa values showed full or almost

Table 4. Adult FHI-I and FHI-S Factor Analysis Eigenvalues

	Eigenvalues			
	1	2	3	4
FHI-I and FHI-S (adulthood items)				
FHI-I (females, $n = 167$)	7.1	2.8	1.9	1.3
FHI-I (males, $n = 158$)	6.6	2.6	1.5	1.2
FHI-S (females, $n = 180$)	5.8	3.7	3.0	2.0
FHI-S (males, $n = 150$)	6.2	2.6	2.3	1.4
FHI-I and FHI-S (childhood items)				
FHI-I (females, $n = 171$)	8.7	6.4	3.5	2.3
FHI-I (males, $n = 160$)	10.0	3.2	2.2	1.6
FHI-S (females, $n = 210$)	7.4	3.9	2.3	1.8
FHI-S (males, $n = 188$)	6.5	3.4	2.3	2.0

perfect agreement. The correlation coefficients between live and consensus sum score for factor1 for each schedule were uniformly high; FHI-I childhood section 0.93, FHI-I adulthood section 0.95; FHI-S childhood section 0.94, FHI-S adulthood section 0.92, (all $P < 0.001$).

Discussion

The Structure and Behavioral Components of the BAP

In this sample of relatives of individuals with ASD, a broad grouping of social-communication impairments, together with rigidity form a consistent BAP trait. Similar factor groupings were found in both childhood and adulthood and in males and females by both self-report and informant interview. The similar structure of child and adult factors and the high child–adult correlations at both the item- and total-score levels suggests considerable developmental stability of BAP traits. The social-communication difficulties identified in this study have also been found by other groups using different measurement methods [Losh, Childress, Lam, & Piven, 2008]. Other quantitative, dimensional measures of the BAP have been developed and are currently in use (for example the Social Responsiveness Scale, Broader Autism Phenotype Questionnaire (BAP-Q), and Broader Autism Phenotype Symptom Scale) [Bernier, Gerdts, Munson, Dawson, & Estes, 2012; Bölte, Poustka, & Constantino, 2008; Sasson et al., 2013]. Piven's findings using the BAP-Q, and our findings, using an interview are similar—both instruments identify the BAP as a broad grouping of social-communication difficulties and rigidity [Sasson et al., 2013]. While studies using two or more BAP measures are required to inform researchers about which instrument characterises the BAP more completely [de Jonge et al., 2014], and whether adding together the FHI-I and FHI-S total scores improves the identification and dimensionalization of the BAP, the major current considerations will rest on whether interview or questionnaire methods are most appropriate for their studies, and whether gathering data from more than one source is important.

Earlier research [Fombonne, Bolton, Prior, Jordan, & Rutter, 1997] suggested that reading and spelling difficulties did not index the BAP unless they formed part of a broader constellation of BAP features. Our findings (Table 2) showed that for both childhood and adult life, there were very low correlations between reading and spelling items and the total score (correlations 0.20–0.30 for childhood and 0.22–0.26 for adult life). Moreover, the inclusion of spelling reduced the Cronbach alpha for the scale as a whole. As shown in Table 5, neither reading nor spelling was included in the factor 1 score. Perfectionism, circumscribed interests, and obsessive/compulsive/ritualistic behaviors, although conceptually associated with factor 1 were excluded from factor 1 on statistical grounds. Accordingly, because factors 1 and 2 scores are so different, we recommend that only factor 1 scores from the FHI-I and FHI-S be used for characterising the core broader phenotype.

Instrument Reliability

The FHI-I and the FHI-S show good overall reliability and identify personality traits and behaviors that persist. Considering the future scoring of the FHIs, the close agreement between live scores and consensus vignette data suggests that live scores are reliable and that the consensus vignette scoring procedure is unnecessary to provide reliable data. However, the blind scoring of some anonymized vignettes is a valuable training experience for interviewers—it is useful to maintain a common FHI calibration for reliability of coding during research, and has considerable utility in research protocols where blindness to other information available to interviewers may need to be preserved.

Convergent Agreement and Validity

Although subject and informant reports were stable over test–retest, particularly at the total-score level, agreement between the two was poor at the level of individual items (see Supporting Information Table A). The two reports may, thus, provide complementary information. The finding of apparent inconsistency in data obtained from different sources is a common feature in ASD and other child and adolescent mental health clinical and research practice [Collishaw, Ford, Rabe-Hesketh, & Pickles, 2009]. Research teams without the resources to administer both interviews could consider obtaining data about multiple relatives (including children under age 11 years) using a single informant. As with ASD, the within-informant stability of the reports for childhood and adulthood are consistent

Table 5. Adult FHI-I[1]/FHI-S[2] Factor Loadings by Item

	Adulthood females ($n = 167^1/n = 180^2$)		Adulthood males ($n = 158^1/n = 150^2$)		Childhood females ($n = 171^1/n = 210^2$)		Childhood males ($n = 160^1/n = 188^2$)	
FHI Item	Item loading factor 1	Item loading factor 2	Item loading factor 1	Item loading factor 2	Item loading factor 1	Item loading factor 2	Item loading factor 1	Item loading factor 2
Reading	0.52/−0.01	0.85/1.27	−0.13/−0.40	1.13/1.19	0.10/0.43	0.68/−0.02	0.26/−0.02	1.00/−0.54
Spelling	0.55/0.14	0.23/0.47	−0.01/−0.79	0.55/0.79	0.88/0.38	0.49/0.05	0.10/0.13	0.99/−0.57
Lack of interest in conversation[a]	0.55/0.83	0.67/−0.37	0.84/0.86	−0.19/−0.21	0.72/0.67	−0.50/−0.36	0.71/0.75	−0.06/−0.06
Reciprocal quality of conversation[a]	0.82/−0.04	−0.16/−0.28	0.74/0.54	0.17/−0.21	0.75/0.19	0.05/−0.40	0.95/0.53	0.07/−0.31
Pragmatics[a]	0.76/0.50	0.07/0.21	0.32/0.15	0.64/0.65	NA	NA	NA	NA
Aloof[a]	0.73/0.85	−0.03/-0.44	0.83/0.99	−0.12/0.26	0.59/0.62	−0.76/−0.55	0.90/0.86	−0.02/−0.11
Friendships[a]	0.66/0.80	−0.09/−0.05	0.73/0.78	−0.01/−0.11	0.22/0.67	−0.36/−0.55	0.74/0.64	−0.11/−0.12
Affection[a]	0.66/0.79	−0.09/−0.16	0.75/0.43	−0.06/0.17	0.79/0.54	−0.24/−0.26	0.67/0.52	−0.20/−0.03
Intimacy[a]	0.76/0.66	0.17/0.30	0.75/0.34	−0.10/0.22	NA	NA	NA	NA
Emotional cues and responsiveness[a]	0.74/0.77	−0.04/0.42	0.84/0.73	0.10/0.06	0.89/0.72	−0.12/−0.14	0.82/0.64	0.19/−0.03
Demonstrativeness[a]	0.64/0.58	0.22/0.07	0.78/0.41	−0.09/0.11	0.47/0.49	−0.63/−0.39	0.83/0.58	0.12/0.07
Social behavior[a]	0.51/0.47	0.25/−0.41	0.46/0.34	0.33/0.30	0.61/0.53	−0.48/−0.23	0.87/0.83	0.28/−0.27
Circumscribed nature of interests	−0.26/0.22	−0.62/0.32	0.46/0.39	0.50/0.03	0.01/−0.09	−1.00/−0.88	0.54/0.21	−0.33/−0.67
Rigidity/openness to experience[a]	0.77/0.61	−0.33/0.02	0.57/0.70	0.20/−0.12	0.62/0.47	−0.20/−0.03	0.64/0.50	0.06/−0.26
Perfectionism	0.18/0.34	−0.49/−0.52	0.21/0.47	0.19/−0.05	NA	NA	NA	NA
Obsessive compulsive/ritualistic	0.38/0.38	−0.88/0.17	0.38/0.36	−0.02/0.37	0.74/0.60	0.05/−0.17	0.65/0.18	−0.83/−0.16

[a] Items selected for factor 1 sum score

with BAP traits appearing early in development and persisting; this is as expected when considering the BAP as one end of an autistic continuum.

Unanswered Questions about the BAP

Despite agreement about the existence of a broader range of social-communication difficulties associated with ASD, many aspects of the BAP that are fundamental to our understanding of ASD and its aetiology remain poorly characterized: The frequency and severity of the BAP in relatives from singleton families, which may be different from multiplex families [Losh et al., 2008], requires further study, as does the relationship between the severities (measured by functional impairment) of BAP traits in parents and their children. Whether traits are familial has methodological implications for researchers utilizing phenotypic information in the search for autism susceptibility genes [Bailey & Parr, 2003]. Importantly, it remains uncertain when (or whether) to classify the relatives as being "affected" with the BAP [Parr, Wittemeyer, & Le Couteur, 2011].

Finally, how best should research groups choose to measure the BAP? The family history method was used in early investigations of the BAP [reviewed by Losh et al., 2011]. In recent years, other groups have conceptualized the BAP as social-communication difficulties at the extreme of a normative trait variation in the general population, and taken a self-report questionnaire approach to measure traits that lead to social impairment [Constantino, 2011; Ronald et al., 2006]. Requiring no researcher visit, this approach enables access to large numbers of relatives and unselected individuals. Whether data gathered using an interview based approach that focuses on specific and pervasive examples of behavior provides higher quality data than that gathered by questionnaire requires investigation.

Future Use of Relatives BAP Status in Genetic Studies

How best can ASD researchers use the phenotypic information gathered with the FHIs in the search for ASD susceptibility genes? The ASD sibling recurrence rate, and frequency of BAP behaviors and traits in relatives are not explained by the rare variants identified to date. Evidence from family studies suggest some unidentified common gene variants (presumably of weak effect) are likely to be important in the aetiology of ASD. In very large genetic studies, collecting BAP data as part of a family assessment will be challenging. However, dimensional BAP data will be important for understanding the mechanisms underpinning ASD, and could be used in analysis of a quantitative trait (QTL) [Pickles et al., submitted 2014]; for the inclusion of relatives affected with the BAP, but not ASD; for phenotypically subtyping family members [Babbs et al., 2014]; and for component scoring of individuals on the two major FHI dimensions. While dimensionalization is often desirable for many genetic approaches, latent class methods may yield a typology useful for formal segregation analyses or sample disaggregation. Future studies of the BAP in relatives from larger autism family samples (such as the Autism Genome Project consortium) are of benefit to enable investigation of "unaffected" parent and sibling BAP status where a proband has a de novo or inherited copy number variant [Pinto et al., 2014]. By categorizing, the BAP status of mothers and fathers, genetic "parent of origin" studies can be undertaken, and subtypes of families can be investigated. New methods for trait analysis combining information from different measures and including different aspects of functioning (such as the range of measures and assessments of the autism/BAP phenotype but also at other levels of investigation including other aspects of behavioral, cognitive styles, ability, and language) will be required.

Strengths and Limitations

This BAP study was undertaken on a large sample of relatives from multiplex autism families likely to show the BAP at high rates. The multicentre nature of the study, with common training and regular consensus reliability meetings, ensured that data were of high quality and findings generalizable. We did not record the interviews undertaken in the UK as during consultation prior to data collection, parents advised us this may affect the quality of the data we received. Parents were concerned that individuals might be less likely to give clear examples about their own behavior or that of their partner if they were being audio or video recorded. The effect of not recording the interviews on the quality of data is unknown. While the revision of the FHI has addressed some previous weaknesses, methodological difficulties remain. First, our interviewers knew that the parents and siblings had a (presumed) genetic liability to ASD and the BAP and were included in the consensus coding process; this knowledge may have influenced the live codings and subsequently, the vignette content and consensus codes. The interview may also be subject to bias owing to relatives of affected pairs being more (or less) sensitive to traits that are similar to those seen in their affected children. Indeed, mothers and fathers may report differently depending on their own BAP status. For this study (as in other BAP studies), mothers usually provided information about themselves, their spouse and their children. Fathers were interviewed about themselves and their spouse only. Finally, as the BAP is found more frequently in males than females

[Pickles et al., 2000], this difference may have contributed to a systematic reporting bias with males less able to identify subtle social-communication behaviors or aspects of rigidity in functioning or alternatively males may have been systematically over-scored.

Considering the general applicability of these data, although the genetic liability in families with two individuals with ASD can be expected to be higher than families in which just one individual is affected [Losh et al., 2008], we would nonetheless expect the form of expression to be similar. Further confirmation through FHI studies of the BAP in relatives of ASD singletons and "high risk siblings" would be desirable; gathering control data from relatives with no family history of ASD will be useful. FHI studies with relatives of children with other neurodevelopmental disorders will also be useful to give further information about the Interviews measurement of the BAP, and behaviors and personality traits caused by other factors [de Jonge et al., 2014].

Conclusion

We have established that the FHI is a reliable dimensional measure of the behavioral BAP with convergent validity across subject and informant forms and is of significant potential value in improving our understanding of the aetiology of ASD. In the future, the FHI will be revised through item reduction, to further enhance its acceptability to researchers and families.

Acknowledgments

We are most grateful to all study participants and to the numerous researchers who contributed to the research. The IMGSAC are grateful for project grants from the following funders: The Medical Research Council, The Wellcome Trust, and The Nancy-Lurie Marks (NLM) Family Foundation (grant GR071470MA); IMGSAC has also received grants from the following funders: BIOMED 2 (CT-97-2759), EC Fifth Framework (QLG2-CT-1999-0094), Telethon—Italy (GGP030227), Fondazione Cassa di Risparmio di Bologna, The Janus Korczak Foundation, Deutsche Forschungsgemeinschaft, Fondation France Telecom, Conseil Regional Midi-Pyrenees, Danish Medical Research Council, Sofiefonden, Swedish Research Council (523-2009-7054), The Beatrice Surovell Haskells Fond for Child Mental Health Research of Copenhagen, and The Danish Natural Science Research Council (9802210).

References

Babbs, C., Lloyd, D., Pagnamenta, A.T., Twigg, S.R., Green, J., McGowan, S.J., et al. (2014). De novo and rare inherited mutations implicate the transcriptional coregulator TCF20/SPBP in autism spectrum disorder. Journal of Medical Genetics, 5, 737–747.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, F.Y., et al. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. Psychological Medicine, 25, 63–77.

Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: The phenotype in relatives. Journal of Autism and Developmental Disorders, 28, 369–392.

Bailey, A., & Parr, J. (2003). Implications of the broader phenotype for concepts of autism. In G. Bock & J. Goode (Eds.), Autism: Neural basis and treatment possibilities Chichester (pp. 26–36). Chichester, UK: Wiley.

Bernier, R., Gerdts, J., Munson, J., Dawson, G., & Estes A. (2012). Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. Autism Research, 5, 13–20.

Bölte, S., Poustka, F., & Constantino J.N. (2008). Assessing autistic traits: Cross-cultural validation of the social responsiveness scale (SRS). Autism Research, 1, 354–363.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., et al. (1994). A case-control family history study of autism. Journal of Child Psychology and Psychiatry, 35, 877–900.

Collishaw, S., Ford, T., Rabe-Hesketh, S., & Pickles, A. (2009). How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general? Journal of Child Psychiatry and Psychology, 50, 571–580.

Constantino, J.N. (2011). Autism as a quantitative trait. In D.G. Amaral, G. Dawson, & D.H. Geschwind (Eds.), Autism spectrum disorders (pp. 510–520). New York, NY: Oxford University Press.

Dawson, G., Webb S., Schellenberg, G.D., Dager, S., Friedman, S., Aylward, E., et al. (2002). Defining the broader phenotype of autism: Genetic, brain, and behavioral perspectives. Developmental Psychopathology, 14, 581–611.

de Jonge, M., Parr, J., Rutter, M., Wallace, S., Kemner, C., Bailey, A., et al. (2014). New interview and observation measures of the broader autism phenotype: Group differentiation. Journal of Autism and Developmental Disorders [Epub ahead of print].

Folstein, S., & Rutter, M. (1977a). Infantile autism: A genetic study of 21 twin pairs. Journal of Child Psychology and Psychiatry, 18, 297–321.

Folstein, S., & Rutter, M. (1977b). Genetic influences and infantile autism. Nature, 265, 726–728.

Fombonne, E., Bolton, P., Prior, J., Jordan, H., & Rutter, M. (1997). A family study of autism: Cognitive patterns and levels in parents and siblings. Journal of Child Psychology and Psychiatry, 38, 667–683.

International Molecular Genetic Study of Autism Consortium. (1998). A full genome screen for autism with evidence for linkage to a region on chromosome 7q. Human Molecular Genetics, 42, 132–137.

International Molecular Genetic Study of Autism Consortium (IMGSAC). (2001). A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. American Journal of Human Genetics, 69, 570–581.

Landa, R., Piven, J., Wzorek, M., Gayle, J.O., Chase, G.A., & Folstein, S.E. (1992). Social language use in parents of autistic individuals. Psychological Medicine, 22, 245–254.

Landa, R., Wzorek, M., Piven, J., Folstein, S.E., & Isaacs, C. (1991). Spontaneous narrative discourse characteristics of parents of autistic individuals. Journal of Speech and Hearing Research, 34, 1339–1345.

Le Couteur, A., Bailey, A.J., Goode, S., Pickles, A., Robertson, S., Gottesman, et al. (1996). A broader phenotype of autism: The clinical spectrum in twins. Journal of Child Psychology & Psychiatry, 37, 785–801.

Lord, C., Rutter, M., DiLavore, P.C., & Risi, S. (2001). Autism diagnostic observation schedule. Los Angeles, CA: Western Psychological Services.

Losh, M., Adolphs, R., & Piven J. (2011). The broad autism phenotype. In D.G. Amaral, G. Dawson, & D.H. Geschwind (Eds.), Autism spectrum disorders (pp. 457–476). New York, NY: Oxford University Press.

Losh, M., Childress, D., Lam, K., & Piven, J. (2008). Defining key features of the broad autism phenotype: A comparison across parents of multiple- and single-incidence autism families. American Journal of Medical Genetetics B Neuropsychiatric Genetics, 147B, 424–433.

Lotter, V. (1967). Epidemiology of autistic conditions in young children: Some characteristics of the parents and children. Social Psychiatry, 1, 163–173.

Muthén, L.K. and Muthén, B.O. (1998–2010). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Parr, J.R., Wittemeyer, K., & Le Couteur, A.S. (2011). The broader autism phenotype: Implications for research and clinical practice. In D.G. Amaral, G. Dawson, & D.H. Geschwind (Eds.), Autism spectrum disorders (pp. 521–526). New York, NY: Oxford University Press.

Pickles, A., Parr, J.R., Rutter, M.L., De Jonge, M.V., Wallace, S., Le Couteur. A.S., et al. (2013). New interview and observation measures of the broader autism phenotype: Impressions of interviewee measure. Journal of Autism and Developmental Disorders, 43, 2082–2089.

Pickles, A., Starr, E., Kazak, S., Bolton, P., Papanikolaou, K., Bailey, A., et al. (2000). Variable expression of the autism broader phenotype: Findings from extended pedigrees. Journal of Child Psychology & Psychiatry, 41, 491–502.

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. American Journal of Human Genetics, 94, 677–694.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). The broader autism phenotype: Evidence from a family study of multiple-incidence autism families. American Journal of Psychiatry, 154, 185–190.

Plomin, R., Haworth, C.M.A., & Davis, O.S.P. (2009). Common disorders are quantitative traits. Nature Reviews Genetics, 10, 872–878.

Ronald, A., Happe, F., Bolton, P., Butcher, L.M., Price, T.S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. Journal of the American Academy of Child and Adolescent Psychiatry, 45, 691–699.

Rutter, M. (2003). Categories, dimensions, and the mental health of children and adolescents. Annals of the New York Academy of Sciences, 1008, 11–21.

Rutter, M. (2009). Understanding and testing risk mechanisms for mental disorders. Journal of Child Psychology & Psychiatry, 50, 44–52.

Rutter, M., Le Couteur, A., & Lord, C. (2003). ADI-R Autism Diagnostic Interview-Revised. Manual. Los Angeles, CA: Western Psychological Services.

Sasson, N.J., Lam, K.S., Childress, D., Parlier, M., Daniels, J.L., & Piven, J. (2013). The broad autism phenotype questionnaire: Prevalence and diagnostic classification. Autism Research, 6(2), 134–143.

Statacorp. (2010) Stata Statistical Software: Release 11. College Station, TX: StataCorp LP..

Szatmari, P., MacLean, J.E., Jones, M.B., Bryson, S.E., Zwaigenbaum, L., Bartolucci, G., et al. (2000). The familial aggregation of the lesser variant in biological and nonbiological relatives of PDD probands: A family history study. Journal of Child Psychology and Psychiatry, 41, 579–586.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. Journal of Autism and Developmental Disorders, 9, 11-29.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table A. Weighted kappa values and percentage agreement between adult FHI-I and FHI-S item scores ($n = 320$).

Table B. Percentage agreement and weighted kappa values for comparison of live and consensus codes for FHI-I and FHI-S childhood and adulthood items.